**Current Issues in Emerging eLearning**

Volume 3
Issue 1 *MOOC Design and Delivery: Opportunities and Challenges*
Article 8

April 2016

# Who is a Student: Completion in Coursera Courses at Duke University

Molly Goldwasser
*Duke University*, molly.goldwasser@duke.edu

Chris Mankoff
*Duke University*, chris.mankoff@icfi.com

Kim Manturuk
*Duke University*, kim.manturuk@duke.edu

Lorrie Schmid
*Duke University*, schmid@duke.edu

Keith E. Whitfield
*Duke University*, keith.whitfield@duke.edu

Follow this and additional works at: https://scholarworks.umb.edu/ciee

Part of the Instructional Media Design Commons, and the Online and Distance Education Commons

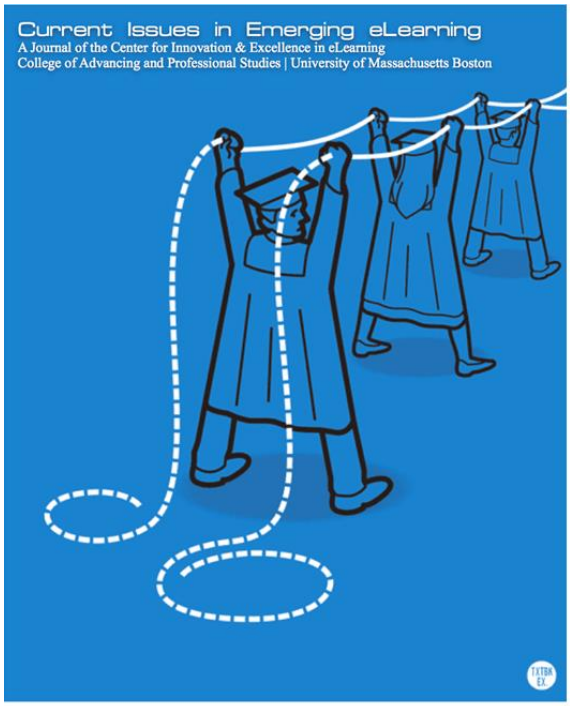
Recommended Citation
Goldwasser, Molly; Mankoff, Chris; Manturuk, Kim; Schmid, Lorrie; and Whitfield, Keith E. (2016) "Who is a Student: Completion in Coursera Courses at Duke University," *Current Issues in Emerging eLearning*: Vol. 3 : Iss. 1 , Article 8.
Available at: https://scholarworks.umb.edu/ciee/vol3/iss1/8

This Article is brought to you for free and open access by ScholarWorks at UMass Boston. It has been accepted for inclusion in Current Issues in Emerging eLearning by an authorized editor of ScholarWorks at UMass Boston. For more information, please contact library.uasc@umb.edu.

# CURRENT ISSUES IN EMERGING ELEARNING

Special Issue on MOOC Design and Delivery: Opportunities and Challenges

Volume 3, Issue 1 (2016/04) ISSN: 2373-6089




EDITOR-IN-CHIEF
Alan Girelli,
University of Massachusetts Boston

ASSOCIATE EDITOR
Apostolos Koutropoulos,
University of Massachusetts Boston

SPECIAL THANK YOU
Leslie P. Limon, copy editor
and revision advisor

JOURNAL COVER IMAGE BY:
Textbook Example,
under Creative Commons licensing.
More work by Textbook example at:
http://textbookexample.com/

COPYRIGHT NOTICE
Current Issues in Emerging eLearning is an
Open Access Journal licensed under a
Creative Commons Attribution -
Non Commercial - No Derivatives 4.0 International License.

Available online at: http://scholarworks.umb.edu/ciee/

# WHO IS A STUDENT: COMPLETION IN COURSERA COURSES AT DUKE UNIVERSITY

Molly Goldwasser, Chris Mankoff, Kim Manturuk,

Lorrie Schmid, and Keith E. Whitfield

Duke University

## ABSTRACT

Much of the interest in MOOCs centers on questions about who completes them. Duke's Coursera-based Massive Open Online Courses (MOOCs) confirm many demographic trends previously delineated by researchers at peer institutions. As found in previous research, this study found individuals who speak English as a first language and who already earned at least a bachelor's degree are the most likely to complete a Coursera course. MOOC researchers to date have not, however, developed clear operational definitions about who constitutes a learner at the outset of the course. This paper proposes some possible definitions to standardize future research. Further, this study looked at factors that predict different learner participation levels and investigated which activities predict Coursera course completion. Study results indicated that viewing online forums and participation in online discussions are both predictive of course completion. The findings suggest that the socio-demographic composition of the group being investigated will depend on how researchers elect to define what a "student" is. Thus, while any of the definitions presented in this paper may be appropriate, depending on what is being studied, the decision of which definition to use should be intentional.

**KEYWORDS**: Massive Open Online Courses (MOOCs), Coursera, Completion, Enrollment, Duke University

# WHO IS A STUDENT: COMPLETION IN COURSERA COURSES AT DUKE UNIVERSITY

Molly Goldwasser[i], Chris Mankoff[ii], Kim Manturuk[iii],

Lorrie Schmid[iv], and Keith E. Whitfield[v]

Duke University

Acknowledgments: Thank you to Dr. Carol Ripple, Associate Director for Education Research and Engagement, and Dr. Lynne O'Brien, Associate Vice Provost for Digital and Online Initiatives for their support of this study.

## INTRODUCTION

Who is a student? In traditional higher education classes, that question is easily answered: Students are people who enroll in a class; if they drop out, they are no longer considered students in the class. However, how do researchers and instructors define who is a student in a massive, open, online class (MOOC)? Unlike students in a traditional college class, students in a MOOC face no consequence for ceasing to participate in a MOOC and have no real incentive to formally withdraw. Similarly, because there is no cost to participate, many people register for a MOOC with no intention of participating throughout the entire course. In this paper, we explore the problem of defining the role of student in a MOOC.

## BACKGROUND

MOOCs have received much publicity in recent years and have become a topic of great interest to researchers. MOOCs are free or very low-cost online courses that typically include instructional videos, assessments, and communication forums; however, new variations on the activities continue to emerge (Beaven, Hauck, Comas-Quinn, Lewis, & de los Arcos, 2014; Fox, 2014). Early research on MOOCs has largely focused on understanding the demographic profile of people who enrolled in courses. For example, an early study looked at data from MIT's first MOOC and found that the people who enrolled were predominately in their 20s and 30s, already had a college degree, and had prior experience in the course topic (Breslow, Pritchard, DeBoer, Stump, Ho, & Seaton, 2013; Emanuel, 2013). Research on other courses and institutions has found similar results (e.g., Christensen, Steinmetz, Alcorn, Bennett, Woods, & Emanuel, 2013; Jordan, 2014).

However, while these studies have documented who enrolls in MOOCs, we believe that there is another question that merits scholarly attention: How do we define the "student" role? There is strong evidence that many people who register for a MOOC have no intention of completing all or any of the activities in

the course (Reich, 2014; Wang & Baker, 2015). Because enrollment has been free, there is no consequence to registering and not participating. Therefore, if researchers use the entire population who register as the basis for their research on course completion, their results are likely to be biased in that it is irrelevant to ask why someone did not finish a course if that person never intended to do so.

We believe the question of who researchers identify as a student is important because much of the discussion around MOOCs has centered on course completion rates. A key criticism of MOOC participation has been the low completion rate among learners (Kolwich, 2013; Yang, Sinha, Adamson, & Rose, 2013). With enrollments well over 10,000 in most courses, completion rates, when calculated as percentages of the original enrollment, are quite low (Catropa, 2013; Jordan, 2014). Kolowich (2013) suggested the overall completion rate of MOOCs hovers around 10%. More recent data suggest that, on average across any MOOC, about 43,000 learners enroll and about 6% complete (Jordan, 2014). However, early MOOC researchers assumed all who registered for a course were students with the potential to complete the course. As one researcher has pointed out, early MOOC learning attracted many people who were "merely curious and tourists from other institutions checking what the fuss was about" (Daniel, 2012).

## SIGNIFICANCE

The concept of providing free college-level courses to the public is not new. As early as the late 1950s, New York University offered two college courses per semester via television through their Sunrise Semester program (Riddle, 2013). Much like MOOCs today, the televised courses enabled students to watch the content for free or to pay a small fee for credit. However, in spite of this history, research on MOOCs is in its infancy and has generally not drawn from prior similar projects. In their review of the published literature between 2008 and 2012, the authors identified only 45 peer-reviewed articles about MOOCs (Liyanagunawardena, Adams, & Williams, 2013). The present analysis represents a significant contribution to this small, yet growing, body of work for three reasons.

First, most prior studies using data from MOOCs have relied on data collected from a single course (e.g., Bell, 2010; Kizilcec, Piech, and Schneider, 2013) One notable exception is a study by Ho et al. (2015), whose research used data aggregated from 18 courses offered by Duke University between 2012 and 2014. This course sample size largely reduced the risk that findings would be biased by unique enrollment patterns in a single course. Second, this paper examines a topic that, to our knowledge, has not been explored in prior research. Many published studies have documented demographic patterns in MOOC enrollment (e.g., Christensen et al., 2013), and researchers have also analyzed the activities people undertake in MOOCs and how those activities relate to course completion (Ho et al., 2015). However, in both of these types of research, the

authors have taken as their total student population the number of people who registered or enrolled in the course. We question this assumption and explore the possible impact of the definition of "student" on research conclusion.

Finally, we relate our analysis to the current debate about the future of MOOCs. Some leaders in the open education movement have been critical of MOOCs because of the low completion rates reported by researchers and universities (Clow, 2013). We contend that this criticism should be reevaluated. Dropout rates in MOOCs are not as high as suggested in prior reports when one controls for intent to complete the course and defines a student as one who is participating in course activity after a pre-determined grace period. Even researchers who do not exclude such people from their counts of students in a MOOC will benefit from some insight regarding how that decision impacts their analyses.

# WHO IS A STUDENT?

MOOC enrollment and persistence statistics consistently classify *completers* as those who have earned some form of a certificate of achievement (in Coursera, these include a Statement of Accomplishment or a Verified Certificate) from the MOOC provider. However, there is no consensus about who constitutes a student at the beginning of the course (DeBoer et al, 2014). Is a student someone who:

- Enrolls in the course?
- Visits the course website?
- Watches a course video?
- Completes an assignment?
- Participates in a discussion forum?
- Some combination of more than one of the criteria listed above?

Traditional education typically waits until the end of a grace period (e.g., drop/add period) to count enrollment and to determine baseline student statistics. If MOOC researchers were to do the same, course completion statistics would increase. However, there is no clear drop point in a MOOC. Some researchers have predicted which students will drop out of a course based on patterns of activity (Halawa, Greene, & Mitchell, 2014) and forum posts (Chaplot, Rhim, & Kim, 2015). These studies focused on predicting dropouts from enrolled and active students. We build on this previous work by assessing who the students are based on the course activities in which they participate. Different demographic groups appear to participate in different course activities; therefore, defining students based on these different participation rates can lead to different research conclusions regarding rates of course completion. In addition, useful information about when and how individuals use course elements, regardless of whether they ultimately complete the course, can inform understandings regarding learner engagement with the material (Kizilcec et al., 2013).

# DATA AND METHODS

In this paper, we present different ways to define a student based on course activities. This includes defining a student as someone who: 1) enrolled in the course, 2) ever visited the course website, 3) watched any video lecture, 4) viewed the discussion forum, or 5) submitted any graded assignment. For each of the five possible definitions, we present regression models that indicate the likelihood of various demographic measures correlating with someone fitting the definition of a student. For example, we find that older course enrollees were more likely to watch any video lecture than younger enrollees. We discuss the implications of these findings for research; how researchers elect to define "students" will impact the socio-demographic composition of the group being investigated. Finally, we present our recommendation that researchers define students as enrollees who attempt at least one graded assessment. We conclude by explaining this recommendation and presenting the next steps for research in this area.

These analyses included all enrolled learners in 18 unique course session offerings comprising 58% of the MOOC offerings at Duke between 2012 and 2014. All courses with complete data were included.[1] See Table 1 for enrollment and activity behaviors (i.e., watching a video, writing a forum post, and receiving a certificate) for each course.

| Course Name / Session | Enrolled | Watched video | Wrote a forum post | Completed assignment | Received SOA or VC |
|---|---|---|---|---|---|
| Bioelectricity / 1 | 18,263 | 7,757 | 814 | 3,727 | 314 |
| Bioelectricity / 2 | 9,795 | 3,956 | 362 | 9,795 | 210 |
| Think Again / 1 | 226,767 | 119,936 | 9,358 | 82,543 | 5,332 |
| Astronomy / 2 | 53,640 | 27,097 | 1,856 | 7,670 | 867 |
| Human Physiology / 1 | 82,437 | 32,583 | 2,185 | 6,665 | 1,036 |
| Human Physiology / 2 | 46,004 | N./A | 1,317 | 3,699 | 871 |
| English Composition / 1 | 82,943 | 36,828 | 11,649 | 3,505 | 1,289 |
| Med Neuroscience / 1 | 66,235 | 21,368 | 2,277 | 12,461 | 590 |
| Med Neuroscience / 2 | 41,985 | 17,668 | 1,184 | 9,855 | 519 |
| Health Innovation / 1 | 43,445 | 11,305 | 2,396 | 4,410 | 3,057 |
| Sports & Society / 1 | 19,394 | 6,073 | 1,092 | 3,402 | 1,629 |
| Sports & Society / 2 | 11,074 | 4,188 | 655 | 1,864 | 1,084 |
| 9/11 & Aftermath / 1 | 16,783 | 6,191 | 911 | 2,648 | 464 |
| Amer Foreign Policy / 1 | 23,720 | 7,850 | 846 | 3,490 | 1,760 |
| Intro to Chemistry / 1 | 34,632 | 14,872 | 1,687 | 8,320 | 556 |
| Higher Education / 1 | 18,809 | 7,247 | 1,311 | 3,679 | 1,532 |
| Marine Megafauna / 1 | 14,374 | 6,989 | 1,305 | 4,232 | 1,469 |
| Data Analysis / 1 | 86,417 | 33,483 | 3,181 | 65,696 | 2,516 |
| Total | 896,717 | 365,391 | 44,386 | 237,661 | 25,556 |

**Table 1:** Duke Coursera Activities by Course

[1] The 42% of courses that were excluded from analysis were omitted due to problems in the source data files, as discussed in the Limitations and Conclusion section of this work

We collected data in two ways: through the Coursera platform and through the use of a pre- and post- survey designed by the Center for Instructional Technology (CIT) at Duke. Demographic indicators used in the analyses include: age, gender, educational level, English as a primary language, race, ethnicity, nationality, and employment status. These were selected because prior research has indicated that these variables correlated with enrollment in and completion of MOOCs (Christensen, 2013; Katy, 2014; Kizilec et al., 2013). We also assessed student activity behaviors, including whether students visited the course website, watched a video, viewed the forum, wrote a forum post, completed a graded assessment, and completed the course. The composite results across all 18 classes on student activities are shown in Table 2.

| Activity | N | % |
|---|---|---|
| Visited course website | 580,664 | 64.75 |
| Watched a video | 365,391 | 40.75 |
| Viewed a forum | 94,232 | 10.51 |
| Wrote a forum post | 44,386 | 4.95 |
| Completed at least one graded | 192,682 | 21.49 |
| Received certification | 25,556 | 2.44 |

**Table 2:** Composite Student Activity Behaviors

Approximately 900,000 learners enrolled in these 18 course session offerings. Fifty-five percent of the learners identified as male, and 45% identified as female. The sample included learners from all over the world and many nationalities. Sixty-three percent identified as White, 22% as Asian, 4% as Black, and 8% as some other category. Sixty-two percent of the sample was aged 34 and younger. Across the whole sample, 35% had completed a bachelor's degree and an additional 30% had advanced degrees. Forty-eight percent reported working full time.

In order to understand how decisions about defining the student body in a MOOC affect subsequent analyses, we began by conducting logistic regression analyses to examine which demographic measures were associated with different criteria for defining students. For example, if we define "students" as those people who ever visited the course website (as opposed to all people who registered), and our models indicate that race is a significant predictor of visiting a course website, then our decision regarding how to define a student will have empirical implications. In the second stage of our analysis, we take course completion as the dependent variable and use both demographic measures and course activity behavior to predict course completion. By comparing which demographic measures were significant in each model, we present a clear example of how research conclusions are affected by how researchers define the student body.

# RESULTS

Tables 3 and 4 present the regression results predicting different categories of student activities. Table 3 presents the results predicting whether someone who enrolled in the course ever visited the course website, ever watched an instructional video, or had ever viewed the discussion forum. People who visited the website, as compared to people who enrolled but never went to the website, were more likely to be male, speak English as their first language, and be aged 35 or older. Learners who participated in watching a video were more likely to identify as Latino or Hispanic and also more likely to be age 35 or older. Those who ever viewed a forum post were more likely to be male, speak English as their first language, and be aged 35 or older. They were also less likely to identify as black or as having already completed college.

| | Visits course website | | Watches a video | | Views the forum posts | |
|---|---|---|---|---|---|---|
| | *β* | *SE* | *β* | *SE* | *β* | *SE* |
| Intercept | 1.65 *** | 0.18 | 0.42 *** | 0.09 | -0.77 *** | 0.09 |
| Male | 0.39 *** | 0.09 | 0.05 | 0.04 | 0.19 *** | 0.04 |
| African American | 0.37 | 0.25 | -0.09 | 0.11 | -0.32 ** | 0.11 |
| Asian | 0.23 | 0.23 | -0.06 | 0.10 | -0.17 | 0.10 |
| Other Races | 0.69 * | 0.28 | 0.18 | 0.11 | -0.13 | 0.10 |
| Hispanic / Latino | -0.12 | 0.12 | 0.16 ** | 0.06 | -0.04 | 0.05 |
| English 1st language | 0.50 *** | 0.09 | 0.06 | 0.05 | 0.12 ** | 0.04 |
| High School or Less | 0.14 | 0.20 | 0.01 | 0.09 | -0.19 * | 0.09 |
| Some College | -0.04 | 0.13 | -0.07 | 0.07 | -0.15 * | 0.06 |
| More than a BA/BS | -0.02 | 0.10 | -0.02 | 0.05 | 0.02 | 0.04 |
| Age – 17 or less | -0.22 | 0.29 | -0.02 | 0.05 | 0.09 | 0.15 |
| Age – 26-34 | 0.10 | 0.12 | 0.18 ** | 0.06 | -0.01 | 0.06 |
| Age – 35-44 | 0.37 ** | 0.14 | 0.38 *** | 0.07 | 0.33 *** | 0.06 |
| Age – 45-54 | 0.70 *** | 0.16 | 0.68 *** | 0.08 | 0.46 *** | 0.07 |
| Age – 55-64 | 1.15 *** | 0.20 | 0.72 *** | 0.09 | 0.67 *** | 0.08 |
| Age – 65 and over | 1.09 *** | 0.25 | 0.87 *** | 0.11 | 0.73 *** | 0.09 |
| *N* | 11295 | | 11295 | | 11295 | |
| Pseudo $R^2$ | 0.0102 | | 0.0170 | | 0.0238 | |

*Note: White, female, BA/BS and 18-25 are the reference groups.*
*Sig p-values are: *<.05, **<.01, ***<.001*

**Table 3:** Regression Models Predicting Passive Course Activity Participation

Table 4 describes findings from our examination of student activity patterns that involve more commitment or effort to complete: writing a forum post, completing an assignment, and/or receiving a certificate. Learners who wrote at least one discussion forum post were more likely to be female and were less likely to have an advanced degree. Given the results of the other models, it is not surprising that people whose first language was English and relatively older learners were more likely to post in the discussion forum.

In an alternate model, we looked at people who completed a course assignment; we found that men, native English speakers, and those older than 35 years old were more likely to complete an assignment. Consistent with other studies, we found that course completion correlated with being a native English speaker, with already having a college degree, and with being aged 35 and older (Christensen et al, 2013).

| | Wrote a forum post | | Completed an assignment | | Received certificate | |
|---|---|---|---|---|---|---|
| | *β* | *SE* | *β* | *SE* | *β* | *SE* |
| Intercept | -1.66 *** | 0.10 | 0.28 * | 0.11 | -2.34 *** | 0.12 |
| Male | -0.24 *** | 0.04 | 0.13 ** | 0.05 | 0.08 | 0.05 |
| African American | 0.13 | 0.12 | -0.15 | 0.13 | -0.29 | 0.15 |
| Asian | 0.08 | 0.11 | -0.06 | 0.13 | 0.01 | 0.13 |
| Other Races | 0.13 | 0.11 | -0.03 | 0.13 | -0.43 *** | 0.15 |
| Hispanic / Latino | 0.03 | 0.06 | 0.01 | 0.07 | -0.10 | 0.07 |
| English 1st language | 0.19 ** | 0.05 | 0.26 *** | 0.05 | 0.22 ** | 0.06 |
| High School or Less | -0.02 | 0.10 | -0.18 | 0.11 | -0.29 * | 0.14 |
| Some College | 0.04 | 0.07 | -0.12 | 0.08 | -0.16 | 0.09 |
| More than a BA/BS | -0.16 ** | 0.05 | -0.06 | 0.06 | 0.20 ** | 0.06 |
| Age – 17 or less | 0.19 | 0.18 | 0.12 | 0.19 | 0.33 | 0.23 |
| Age – 26-34 | 0.26 ** | 0.07 | 0.03 | 0.08 | 0.11 | 0.09 |
| Age – 35-44 | 0.41 *** | 0.07 | 0.08 | 0.08 | 0.38 *** | 0.09 |
| Age – 45-54 | 0.64 *** | 0.08 | 0.30 ** | 0.09 | 0.46 *** | 0.09 |
| Age – 55-64 | 0.62 *** | 0.09 | 0.29 ** | 0.10 | 0.46 *** | 0.10 |
| Age – 65 and over | 0.47 *** | 0.11 | -0.05 | 0.11 | 0.10 | 0.13 |
| *N* | 11295 | | 7929 | | 11295 | |
| Pseudo $R^2$ | 0.0104 | | 0.0076 | | 0.0100 | |

*Note: White, female, BA/BS and 18-25 are the reference groups.*
*Sig p-values are: *<.05, **<.01, ***<.001*

**Table 4:** Regression Models Predicting Active Course Activity Participation

These findings highlight the need to make intentional and research-driven decisions about defining a student in a MOOC. Depending on the criteria used to define a student, we may find, for example, that students in a course are more likely to be male or to have an advanced degree. We continued to illustrate this point in the second set of analyses by conducting two sets of logistic regressions predicting course completion, focusing on participation in the forums. In one case we defined as students the participants who had viewed discussion posts (yielding findings represented in Table 5). In another case we defined as students those who posted on a forum site (yielding findings represented in Table 6). Two models were conducted for each regression. Model 1 includes only the forum indicator of interest, and Model 2 includes the indicator as well as demographic variables.

As seen by comparing the two analyses, the model including the variable for viewing the forum generates a significant negative coefficient for the Hispanic/Latino variable. However, the same measure is not significant in the model including the variable indicating someone had posted in the forum. This illustrates how research decisions regarding what course activities qualify someone as a student affect the results of an analysis of course completion.

| | Model 1 | | Model 2 | |
|---|---|---|---|---|
| | *β* | *SE* | *β* | *SE* |
| Intercept | -4.75 *** | 0.01 | -3.20 *** | 0.06 |
| Viewed Forum | 2.98 *** | 0.02 | 2.20 *** | 0.03 |
| Male | | | -0.33 *** | 0.03 |
| African-American | | | -0.77 *** | 0.17 |
| Asian | | | -0.29 *** | 0.08 |
| Other races | | | -0.90 *** | 0.03 |
| Hispanic/Latino | | | -0.18 ** | 0.07 |
| English 1st language | | | -0.45 *** | 0.03 |
| Age | | | 0.13 *** | 0.01 |
| *N* | 896,717 | | 110,206 | |
| Pseudo $R^2$ | 0.20 | | 0.15 | |

*Note: White and female are the reference groups.*
*Sig p-values are: *<.05, **<.01, ***<.001*

**Table 5:** Predicting Course Completion from Viewing the Forum Postings

| | Model 1 | | Model 2 | |
|---|---|---|---|---|
| | *β* | *SE* | *β* | *SE* |
| Intercept | -4.29 *** | 0.02 | -3.00 *** | 0.06 |
| Posted in forum | 2.95 *** | 0.01 | 1.89 *** | 0.03 |
| Male | | | -0.52 *** | 0.03 |
| African-American | | | -0.97 *** | 0.17 |
| Asian | | | -0.34 *** | 0.08 |
| Other Races | | | -0.94 *** | 0.03 |
| Hispanic / Latino | | | -0.10 | 0.07 |
| English 1st language | | | -0.08 ** | 0.03 |
| Age | | | 0.11 *** | 0.01 |
| N | 896,717 | | 110,206 | |
| *Pseudo* $R^2$ | 0.15 | | 0.13 | |

*Note: White and female are the reference groups.*
*Sig p-values are: *<.05, **<.01, ***<.001*

**Table 6:** Predicting Course Completion from Writing Forum Postings

# DISCUSSION AND NEXT STEPS

The findings of the current study highlight the importance of defining who is a student when looking at patterns of participation and completion in MOOCs. Important in these findings is that education, age, and gender matter in distinctive ways depending on how one defines the population of interest. Our results suggest that older learners, while a smaller proportion of the overall population of MOOC learners, are more likely to watch a video but less likely to complete the course than younger participants. These differences may indicate that learners of different ages may have different intentions when registering for a MOOC. It may also reflect generational differences in the way learners consume information. It may be that younger adults are used to searching for bits of information from multiple sources and use multiple resources to obtain knowledge. Older adults on the other hand may be using traditional approaches to knowledge acquisition.

Also interesting were the gender-based findings. While more men enrolled than women, women were more likely to engage with the course by writing a forum post. There has been much discussion of gender differences in the style and content of computer- mediated communication (e.g., Herring, 2000). Many instructors of MOOCs are interested in the utility of the forums for discussing course material and creating community among geographically diverse course participants. Our results indicate that, while most learners do not

participate in the discussion forums, those who do are more likely to complete the course.

Demographic variables in this study were defined by traditional U.S. American classifications. Additional research is needed to examine student trends by sub-category according to different global norms. There is also a need for content analyses of the posts to see if there are gender differences. Future research is also needed to investigate how lessons learned from MOOCs impact traditional students on campus.

## LIMITATIONS AND CONCLUSION

The data used for this research have some limitations. Almost half of the data files we obtained had errors that made them unusable in this analysis. These tended to be the data files generated in courses run relatively early in the history of use of the MOOC platform, so our analyses may not be as applicable to MOOCs offered early in the project. The most serious limitation in this study, and one that often affects research on MOOCs, involves selection bias. The large numbers of people who enrolled yet never participated in any course activities were also people who were less likely to complete the demographic survey or the pre- and post- surveys. In future research, we hope to use analytic techniques to account at least partially for selection bias; however, that was not possible with this project. We therefore offer the caveat that the analysis presented here should be taken as illustrative of the need to make theoretically-based decisions about defining who a student is, while acknowledging that the empirical findings related to predicting course activities may not generalize to other courses.

In conclusion, we recommend that researchers define a student based on the research question under investigation. When looking at completion rates, as many recent studies have done, it logically follows to consider a student to be anyone who has attempted at least one assessment. These are the people enrolled in the course who are most likely to intend to complete the course. This definition excludes people who enrolled simply to watch videos or explore the course structure. Alternatively, if researchers are interested in analyzing patterns of movement in a course—the order in which people move through materials—it logically follows that they would want to include all participants who ever visited the course website. Any of the definitions of who is a student presented in this paper may be appropriate depending on what is being studied, but the decision of which definition to use should be one made intentionally and not by default, as has often been done to date.

# REFERENCES


Beaven, T., Hauck, M., Comas-Quinn, A., Lewis, T., & de los Arcos, B. (2014). MOOCs: Striking the right balance between facilitation and self-determination. *MERLOT Journal of Online Learning and Teaching*, *10*, 31-43.

Bell, F. (2011). Connectivism: Its place in theory-informed research and innovation in technology-enabled learning. *The International Review of Research in Open and Distance Learning, 12*(3): 98-118.

Breslow, L., Pritchard, D., DeBoer, J., Stump, G., Ho, A., & Seaton, D. (2013). Studying learning in the worldwide classroom: Research into edX's first MOOC", *Research and Practice in Assessment*, *8*, 13-25.

Catropa, D. (2013). Big (MOOC) data. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/blogs/stratedgy/big-mooc-data.

Chaplot, D. S., Rhim, E., & Kim, J. (2015). Predicting student attrition in MOOCs using sentiment analysis and neural networks. In *AIED 2015 Workshop Proceedings – Vol 3*. Retrieved from http://ceur-ws.org/Vol-1432/islg_pap2.pdf

Christensen, G., Steinmetz, A., Alcorn, B., Bennett, A., Woods, D., & Emanuel, E. J. (2013). The MOOC phenomenon: Who takes massive open online courses and why? Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2350964

Clow, D. (2013). MOOCs and the funnel of participation. *In Third International Conference on Learning Analytics and Knowledge (LAK 2013)*. Leuven, Belgium. doi: 10.1145/2460296.2460332_.

Daniel, J. (2012). Making sense of MOOCs: Musings in a maze of myth, paradox, and possibility. *Journal of Interactive Media in Education*.

DeBoer, J., Ho, A. D., Stump, G. S., & Breslow, L. (2014). Changing "course" Reconceptualizing educational variables for massive open online courses. *Educational Researcher*, *43*, 74-84. doi: 0013189X14523038.

Emanuel, E. (2013). Online education: MOOCs taken by educated few. *Nature*, 503 (7476), 342. doi: 10.1038/503342

Fox, A. (2014). From MOOCs to SPOCs. *Communications of the ACM, 56* (12), 38-40. doi: 10.1145/2355.918

Halawa, S., Greene, D., & Mitchell, J. (2014). Dropout prediction in MOOCs using learner activity features. *Proceedings of the European MOOC Summit.* Lausanne, Switzerland.

Herring, S.C. (2000). Gender differences in CMC: Findings and implications. *Computer professionals for social responsibility newsletter, 18*(1). Retrieved from http://cpsr.org/issues/womenintech/herring/.

Ho, A. D., Chuang, I., Reich, J., Coleman, C. A.,Whitehill, J., Northcutt, C. G., Williams, J. J., et al. (2015). HarvardX and MITx: Two years of open online courses Fall 2012 – Summer 2014. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2586847.

Jordan, Katy. (2014). Initial trends in enrollment and completion of massive open online courses. *The International Review of Research in Open and Distributed Learning*, *15*(1).

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In D. Suthers, K. Verbert, E. Duval, & X. Ochoa (eds.), *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (pp. 170-179). Leuven: Solar

Kolowich, S. (2013). Coursera takes a nuanced view of MOOC dropout rates. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/blogs/wiredcampus/coursera-takes-a-nuanced-view-of-mooc-dropout-rates/43341

Liyanagunawardena, T. R., Adams, A. A., & Williams, S. A. MOOCs: A systematic study of the published literature, 2008-2012. *Review of Research in Open and Distance Learning*, *14*(3), 202-227.

Riddle, Randy. (2013). MOOC Pre-History. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/blogs/technology-and-learning/mooc-pre-history.

Reich, J. (2014). MOOC completion and retention in the context of student intent. *Educause Review*. Retrieved from http://www.educause.edu/ero/article/mooc-completion-and-retention-context-student-intent

Seaton, D. T., Bergner, Y., Chuang, I., Mitros, P., & Pritchard, D. E. (2014). Who does what in a massive open online course? *Communications of the ACM*, *57*(4), 58-65.

Wang, Y., & Baker, R. (2015). Content or platform: Why do students complete MOOCs? *MERLOT Journal of Online Learning and Teaching*, *11*(1), 17-30.

Yang, D., Sinha, T., Adamson, D., & Rose, C. P. (2013). Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *Proceedings of the 2013 NIPS Data-Driven Education Workshop* (Vol. 11, p. 14).

# AUTHOR BIOS

[i] **Molly Goldwasser** is the Associate Vice Provost for Academic Affairs at Duke University. At the time of article submission, she served as the Assistant Vice Provost and Manager of Institutional Assessment and Accreditation in the Office of the Provost. Her professional background and her research portfolio center on assessment, educational access and equity, and teaching and learning. Goldwasser earned a Bachelor's degree in Public Policy Studies from Duke University, a Master's degree in Curriculum and Instruction from the University of Mississippi, and a Doctorate of Education in Education Leadership and Policy from Vanderbilt's Peabody College. Email: molly.goldwasser@duke.edu

[ii] **Chris Mankoff** is an Associate at ICF International. In this position he programs and processes commercial and government survey data. He has extensive experience managing educational data, and has published and presented on online education. He formerly served as a data manager/application developer for the Duke Center for Autism and Brain Development. In this role he managed vendors in the development of clinical trials database applications, developed clinic trial datasets for use by statisticians, and designed web-based data capture forms. Prior to January 2015, he served as the online data manager for Duke University, in which position he collected, managed, distributed and analyzed Coursera data for classes taught by Duke University professors. Email: chris.mankoff@icfi.com

[iii] **Kim Manturuk** manages research and evaluation for the Center for Instructional Technology at Duke University. She is responsible for data collection and analysis to assess the effectiveness of Duke's online course programs. She also collaborates with faculty teams to evaluate experiments and innovations in educational technologies. Kim is particularly interested in how online technologies can expand access to education for people with disabilities. Kim holds an M.A. and Ph.D. in Sociology from the University of North Carolina at Chapel Hill and received her B.A. from Indiana University.

[iv] **Lorrie Schmid, MA, PhD** is a recent graduate of the Department of Educational Psychology at the School of Education at the University of North Carolina, Chapel Hill. Her work as manager, data and analytics for the Education and Human Development (EHD) Incubator) involves supporting research and research collaborations related to data on education and human development through providing data management expertise to EHD researchers focusing on capacity to store, analyze and link sensitive data. Her interests include: adolescent social development, and issues of peer influence, as well as methodological and statistical interests in data matching, secure brokering, and data harmonization. She previously served as a Data Manager and Research Analyst at the Center for Child and Family Policy, Duke University. Email: schmid@duke.edu

[v] **Keith E. Whitfield** is the Vice Provost for Academic Affairs at Duke University. He holds appointments as Professor in the Department of Psychology and Neuroscience and Research Professor in the Department of Geriatric Medicine at Duke University Medical Center. He is also the co-Director of Duke's Center on Biobehavioral Health Disparities Research. He earned a Bachelor's degree in Psychology from the College of Santa Fe, Santa Fe, NM and a M.A. and Ph.D. in Life-Span Developmental Psychology from Texas Tech University in Lubbock, TX. He also received post-doctoral training in quantitative genetics from the University of Colorado in Boulder, CO. He has published more than 180 articles, books and book chapters on cognition, health, and individual development and aging with a focus on African Americans. He currently serves as the Managing Editor for the journal *Ethnicity and Health* and has served as an ad-hoc reviewer and/or editorial board member of multiple journals. He is the member of several professional associations and a past chair of the Behavioral and Social Sciences Section of the Gerontological Society of America.